(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 October 2001 (11.10.2001)

(10) International Publication Number
PCT WO 01/74298 A2

(51) International Patent Classification⁷: A61K

(21) International Application Number: PCT/US01/10661

(22) International Filing Date: 2 April 2001 (02.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/193,614 31 March 2000 (31.03.2000) US

(71) Applicants (for all designated States except US): BROWN UNIVERSITY RESEACH FOUNDATION [US/US]; 42 Charlesfield Street, Providence, RI 02912 (US). HOWARD HUGHES MEDICAL INSTITUTE [US/US]; 4000 Jones Bridge Road, Chevy Chase, MD 20815 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ALBERINI, Cristina, M. [US/US]; 550 Wayland Way, Providence, RI 02912 (US). BEAR, Mark, F. [US/US]; 244 Metacom Avenue, Bristol, RI 02806 (US).

(74) Agents: VINCENT, Matthew, P. et al.; Ropes & Gray, One International Place, Boston, MA 02110-2624 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR REGULATING MEMORY CONSOLIDATION

(57) Abstract: The present invention is based on discovery of genes that are up- or down-regulated in inhibitory avoidance, e.g., long-term memory, which genes are therefor believed to have roles in memory consolidation. In particular, we have discovered that memory consolidation involves the regulation of expression of such genes as zif268 (EGR1), insulin-like growth factor (IGF-1), glutamate receptor (1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF. For ease of reading, these genes are collectively referred to herein as "LTM genes", and their encoded proteins as "LTM proteins".





newly-synthesized proteins are additional transcription factors that ultimately give rise to the activation of late response genes, whose products are responsible for the modification of synaptic efficacy leading to LTM.

CREB subserves the formation of memories of various types of tasks that utilize different brain structures. Evidence is available suggesting that CREB regulates the transcription of genes that subserve LTM. In Aplysia, for example, CREB activation has been interfered with by microinjection of CRE containing oligonucleotides into cultured neurons. In Drosophila, CREB function has been disrupted using a reverse genetic approach. Thus, LTM has been specifically blocked by the induced expression of a CREB repressor isoform, and enhanced by the induced expression of an activator isoform. In mouse, the role of CREB has been confirmed by behavioural analy[redacted] knock-out line with a targeted mutation in the CREB gene. In these mutants, le[redacted] and short term memory are normal, whereas long term memory is impaired. On the whole, the data suggest that encoding of long term memories involve highly conserved molecular mechanisms.

Animals with lesions of the medial temporal lobes and related thalamic structures show a profound disruption of memory consolidation. We have previously demonstrated that fornix-dependent lesion-induced amnesia is associated with abnormal regulation of gene expression in specific subregions of the hippocampus. See, for example, Taubenfeld et al. (1999) Nat Neurosci 2:309-10. In normal animals, inhibitory avoidance training produces a rapid and persistent increase in the phosphorylation of CREB, which is a necessary step in the regulation of CRE-mediated gene expression required for memory consolidation. The change in CREB phosphorylation is largely confined to hippocampal fields CA1 and dentate gyrus, and lasts at least 6 hours after training. Animals with fornix lesions learn the inhibitory avoidance and display memory at control levels for up to 6 hours, however, by 24 hours they exhibit amnesia. The amnesic animals also fail to exhibit any increase in hippocampal CREB phosphorylation after training. Our results suggest that hippocampal inputs passing through the fo[redacted] regulate consolidation of this form of memory via regulation of CREB-mediated [redacted] expression in hippocampal neurons.

Initial learning is likely to result from changes in the transmission of synapses conveying information about where the animal is in space. Whether or not these

growth factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF;

(ii) contacting said system with a test compound; and

(iii) determining if the test compound alters the level of expression of the gene.

In certain preferred embodiments, the reaction system is a cell-free system, such as a purified protein preparation or a cell-lysate. In other embodiments, the reaction system is a whole cell system.

In preferred embodiments, the assay can be used to identify agents which modulate memory consolidation from amongst a plurality of different test agents.

In certain preferred embodiments, the test compound can be small organic molecules, e.g., those having a molecular weight less than 2500 amu.

Still another aspect of the present invention provides a method of conducting a drug discovery business comprising:

(i) identifying, by one or more of the above drug discovery assay, a test compound which the level of expression of the gene or the activity of the gene product;

(ii) conducting therapeutic profiling of agents identified in step (i), or further analogs thereof, for efficacy and toxicity in animals; and

(iii) formulating a pharmaceutical preparation including one or more agents identified in step (ii) as having an acceptable therapeutic profile.

In certain preferred embodiments, the business method includes an additional step of establishing a distribution system for distributing the pharmaceutical preparation for sale, and may optionally include establishing a sales group for marketing the pharmaceutical preparation.

Yet another aspect of the present invention provides a method of conducting a target discovery business comprising:

(i) identifying, by one or more of the above drug discovery assay, a test compound which the level of expression of the gene or the activity of the gene product;

(ii) (optionally) conducting therapeutic profiling of agents identified in step (i), or further analogs thereof, for efficacy and toxicity in animals; and

(iii). licensing, to a third party, the rights for further drug development of said identified agents.





Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

**Brief Description of the Drawings**

Figure 1. Time-course Northern blot analysis of zif268 and c-fos following IA training. Increase in zif268, but not c-fos mRNA is evident in all animals at 9 and 20 hr after training.

Figure 2. Time course Northern blot analysis of C/EBPβ and cyclophilin (control) mRNA following IA training in hippocampi of unoperated and fornix-lesion rats.

Figure 3. Broad scale expression profiling with cDNA expression arrays. Side by side hybridizations with cDNA probes prepared from two different RNA populations allow the simultaneous comparison of the expression levels of all the cDNAs on the array. (from Clontech user manual).

Figure 4. Examples of changes in gene array hybridizations reflecting differential expression of mRNAs following IA training. Hippocampi of control rats are compared to hippocampi of rats trained and sacrificed 9 hr later. Note that on these arrays each sequence is spotted in duplicate.

Figure 5. A Northern blot test to confirm the levels of certain transcripts.

Figure 6. Panel A. Mean escape latency for rats trained on the water maze. Rats received eight trials a day, for four consecutive days. Panel B. Mean escape latency on each of the eight trials of day one (trials 1-8) and day two (trials 9-16).

**Detailed Description of the Invention**

I. Overview

The present invention is based on the discovery of genes that are up- or down-regulated in inhibitory avoidance, e.g., long-term memory, which genes are therefor believed to have roles in memory consolidation. In particular, we have discovered that memory consolidation involves the regulation of expression of such genes as zif268 (EGR1), insulin-like growth factor (IGF-1), glutamate receptor 1 (GluR1), glutamate

These basic-region, leucine-zipper proteins bind to DNA sequences, called cAMP response element (CRE) sites, which are often found in the upstream regulatory regions of genes whose synthesis is cAMP responsive. Molecular analysis has shown that CRE sites, and their interaction with CREB family members, are necessary for cAMP responsiveness. After the catalytic subunit of PKA translocates to the nucleus, it can directly phosphorylate the serine residue at position 133 on CREB, thus activating the protein and directly linking the cAMP transduction pathway to the induction of new gene expression (Backsai et al. (1993) Science 260: 222–226; and Hagiwara et al. (1993) Mol Cell Biol 13:4852–4859). CREB is also phosphorylated via other kinases, such as described above and in Deisseroth et al. (1996) Neuron 16:89-101; Impey et al. (1996) Neuron 16:973-82; and Impey et al. (1998) Neuron 21:869-883.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding" a polypeptide may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid molecule comprising an open reading frame encoding a polypeptide, including both exon and (optionally) intron sequences. The term "intron" refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

"Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. The term "percent identical" refers to sequence identity





are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. The term "modulation" as used herein refers to both upregulation, i.e., stimulation, and downregulation, i.e. suppression, of a response.

The "non-human animals" of the invention include mammalians such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that one of the recombinant genes is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "promoter" means a DNA sequence that regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in cells. The term encompasses "tissue specific" promoters, i.e. promoters, which effect expression of the selected DNA sequence only in specific cells (e.g. cells of a specific tissue). The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well. The term also encompasses non-tissue specific promoters and promoters that constitutively express or that are inducible (i.e. expression levels can be controlled).

The terms "protein", "polypeptide" and "peptide" are used interchangably when referring to a gene product.

The term "recombinant protein" refers to a polypeptide of the present invention which is produced by recombinant DNA techniques, wherein generally, DNA encoding a polypeptide is inserted into a suitable expression vector which is in turn used to

and any other nucleic acid, (e.g. as intron), that may be necessary for optimal expression of a selected nucleic acid.

A "transgenic animal" refers to any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of a proteins, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption is caused by human intervention, including both recombination and antisense techniques.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forts of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

III. Exemplary Embodiments





be varied, or temperature of salt concentration may be held constant while the other variable is changed. Preferred nucleic acids have a sequence at least 75% homologous and more preferably 80% and even more preferably at least 85% homologous with an nucleic acid sequence of an LTM gene. Nucleic acids at least 90%, more preferably 95%, and most preferably at least about 98-99% homologous with a nucleic sequence of an LTM gene are of course also within the scope of the invention.

Nucleic acids having a sequence that differs from the nucleotide sequences shown in one of SEQ ID NOs: 1-X due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a LTM polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect the amino acid sequence of a LTM polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject LTM polypeptides will exist among mammalians. One skilled in the art will appreciate that these variations in one or more nucleotides (e.g., up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a LTM polypeptide may exist among individuals of a given species due to natural allelic variation.

*B. LTM proteins*

Certain assays of the present invention use isolated or recombinant LTM polypeptides which are isolated from, or otherwise substantially free of other cellular proteins, especially other signal transduction factors and/or transcription factors which may normally be associated with the LTM polypeptide. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of LTM polypeptides having less than about 20% (by dry weight) contaminating protein, and preferably having less than about 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial

the present invention is a mammalian LTM protein. It will be understood that certain post-translational modifications, e.g., phosphorylation and the like, can increase the apparent molecular weight of the LTM protein relative to the unmodified polypeptide chain.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of one of the subject LTM polypeptides which function in a limited capacity as one of either an LTM agonist (mimetic) or an LTM antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of naturally occurring forms of LTM proteins.

Homologs of each of the subject LTM proteins can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the LTM polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to a downstream or upstream member of the LTM cascade which includes the LTM protein. In addition, agonistic forms of the protein may be generated which are constitutively active. Thus, the LTM protein and homologs thereof provided by the subject invention may be either positive or negative regulators of memory consolidation.

*C. Cells expressing LTM proteins*

As described below, the assays of the invention may include cells transfected to express a recombinant form of the subject LTM polypeptides. The host cell may be any prokaryotic or eukaryotic cell. Thus, a nucleotide sequence derived from the cloning of mammalian LTM proteins, encoding all or a selected portion of the full-length protein, can be used to produce a recombinant form of an LTM polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. MAP kinase,





a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the in vitro system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with upstream or downstream elements or with intrinsic enzymatic activity. Many of the LTM proteins identified by the subject method will be amenable to some form of cell-free assay formats. Soluble proteins, be they cytoplasmic or extracellular, can be recombinantly expressed and at least partially purified, or provided as lysates, for use in cell-free assays. Membrane-associated proteins can, in certain instances, be purified in detergent or liposomes, or isolated as part of a cell membrane fraction or organelle preparation.

Accordingly, in an exemplary screening assay of the present invention, [redacted]on mixture is generated including the LTM protein and one or more proteins (or nucleic acids) which interact with the LTM protein, such molecules being referred to herein as "LTM-interacting partners" or "LTM-IP". Examples of LTM-IP include proteins that function upstream (including both activators and repressors of LTM activity), and proteins or nucleic acids which function downstream of the LTM polypeptide, whether they are positively or negatively regulated by it. The reaction mixture also includes one or more test compounds. Detection and quantification of complexes of the LTM protein with upstream or downstream LTM-IP provide a means for determining a compound's efficacy at inhibiting or potentiating complex formation between LTM and the LTM-IPs. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In one control assay, isolated and purified LTM polypeptide is added to a composition containing the LTM-IP, and the formation of a complex is quantitated in the absence of the test compound.

Complex formation between the LTM polypeptide and a binding partner may be detected by a variety of techniques. Modulation of the formation of complexes [redacted] quantitated using, for example: detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled proteins; by immunoassay; or by chromatographic detection.

immunodetection of complexes using antibodies reactive with the LTM binding partner, or which are reactive with the LTM protein and compete with the binding partner; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the binding partner, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with an LTM-IP. To illustrate, the LTM-IP can be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine tetrahydrochloride or 4-chloro-1-napthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J. Biol. Chem. 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as anti-LTM antibodies, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes a second polypeptide sequence for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include mycepitopes (e.g., see Ellison et al. (1991) J. Biol. Chem. 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, N.J.).

Other cell-free ebodiments include assays which detect an intrinsic activity of an LTM protein or a complex including an LTM protein, and identify compounds that increase or inhibit that activity. For instance, the reaction micture can be generated to include the LTP protein, a substrate for an enzymatic activity of the LTM protein, and the test agent. The rate of conversion of the substrate to product is determined, and can be compared to such control samples as the LTM proteins and substrate admixed alone. Test agents which are inhibitors of the LTM activity will decrease the rate of conversion of the substrate to product, whereas test agents that increase that rate are likely to be agonists of the LTM activity.





In yet another aspect of the invention, the subject drug screening assays can utilized the LTM proteins to generate a "two hybrid" assay (see, for example, U.S. Pat. No. *5,283,317*; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and PCT Publication No. WO 94/10300). Briefly, the two hybrid assay relies on reconstituting in vivo a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first chimeric gene can be generated with the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for an LTM protein. The second hybrid protein encodes a transcriptional activation domain fused in frame to another polypeptide, e.g., and LTM-IP, which binds to the LTM protein. If the two fusion proteins are able to interact, e.g., form an LTM-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site which is bound by the DNA-binding domain of the first fusion proteins, and expression of the reporter gene can be detected and used to score for the interaction of the LTM and sample proteins.

a. Exemplary LTM proteins: GluR1 and GluR2

In certain embodiments, the subject assays are used to identify compouds which modulate the activity of a glutamate receptor, such as GluR1 (SEQ ID Nos. 1 and 2) or GluR2 (SEQ ID Nos. 3 and 4).

The subject assays can be used to identifying agents which bind to the receptors and, e.g., mimic or potentiate the activity of the natural ligand, or which inhibit binding or signal transduction by the receptor. To illustate, for binding studies, test agents can be tested for competition with binding. For example, [$^3$H]AMPA binding can be assessed as follows: cells expressing GluR1 or GluR2 are preincubated in 50 mM Tris-HCl buffer, pH 7.4, for 30 min, and then incubated at 4°C for 4h in 100 mM buffer plus 100 mM KCSN, 70 nM [$^3$H]AMPA (53 Ci/mmol, NEN, Boston, MA) and the test agent. Binding of the radiolabeled AMPA is assessed and compared to the level of binding in the absence of the test compound. Nonspecific binding is measured in the presence of 1 mM L-glutamate.

c. Exemplary LTM proteins: neuroendocrine VGF

The neurotropin-inducible gene VGF (SEQ ID Nos. 11 and 12) is expressed in neuronal and endocrine tissues. It encodes a secretory protein that is proteolytically processed in neuronal cells to low molecular mass polypeptides. In addition to targeting the second messenger induction caused by contacting cells with VGF, the subject assays can also be used to identify agents which inhibit the proteolytic processing of VGF.

d. Exemplary LTM proteins: ZIF268

Zif268 (SEQ ID Nos. 5 and 6) is also known as krox-24, egr-1, TIS 8, NGFI-A or zenk. It is a zinc-finger transcription factor which binds to a so-called "EGR1 motif", e.g., a trascriptional regulatory sequence of 5'-CGCCCCCGC or 5'-GCGTGGGCG. See, for example, Rauscher et al. (1990) Science 250:1259; and Pavletich et al. (1991) Science 252:809.

In certain embodiments, the subject assay can be a transcription based assay including a reporter gene having an EGR1 motif as part of its transcriptional regulatory sequences. Test agents can be assessed for their ability to enhance or inhibit Zif268-dependent transcription.

In other embodiments, test agents can be tested for their ability to ehance or inhibit binding of Zif268 with its EGR1 motif in a competitive binding assay, e.g, a nucleic acid including the motif.

e. Exemplary LTM proteins: C/EBPβ

In another embodiment, the target for the subject drug screening assay is the transcription factor C/EBPβ (SEQ ID Nos. 9 and 10), also referred to as the CCAAT/enhancer protein. The CCAAT/Enhancer Binding Protein (C/EBP) family belongs to the basic leucine zipper class of transcription factors. The C/EBP protein binds to the CCAAT-box (consensus $GG^{T}/_{C}CAATCT$).

As above with Zif268, C/EBP-dependent transcriptional activity or competition binding assays can be used to assessed the ability of test compounds to enhance or inhibit C/EBPβ activity.

f. Exemplary Screening and Selection Assays: Second Messenger Generation

When screening for bioactivity of test compounds, intracellular second messenger generation can be measured directly. A variety of intracellular effectors have been identified as being regulated by certain of the LTM proteins described above,





kinase, but activates protein kinases or phosphatase that function downstream in the signal transduction pathway.

One such cascade is the MAP kinase pathway that appears to mediate both mitogenic, differentiation and stress responses in different cell types. Stimulation of growth factor receptors results in Ras activation followed by the sequential activation of c-Raf, MEK, and p44 and p42 MAP kinases (ERK1 and ERK2). Activated MAP kinase then phosphorylates many key regulatory proteins, including p90RSK and Elk-1 that are phosphorylated when MAP kinase translocates to the nucleus. Homologous pathways exist in mammalian and yeast cells. For instance, an essential part of the S. cerevisiae pheromone signaling pathway is comprised of a protein kinase cascade composed of the products of the STE11, STE7, and FUS3/KSS1 genes (the latter pair are distinct, functionally redundant). Accordingly, phosphorylation and/or activation of members of this kinase cascade can be detected and used to quantitate receptor engagement. Phosphotyrosine specific antibodies are available to measure increases in tyrosine phosphorylation and phospho-specific antibodies are commercially available (New England Biolabs, Beverly, MA).

In yet another embodiment, the signal transduction pathway of the LTM protein upregulates expression or otherwise activates an enzyme which is capable of cleaving a substrate which can be added to the cell. The signal can be detected by using a detectable substrate, in which case lose of the substrate signal is monitored, or alternatively, by using a substrate which produces a detectable product. In preferred embodiments, the conversion of the substrate to product by the activated enzyme produces a detectable change in optical characteristics of the test cell, e.g., the substrate and/or product is chromogenically or fluorogenically active. In an illustrative embodiment the, signal transduction pathway causes a change in the activity of a proteolytic enzyme, altering the rate at which it cleaves a substrate peptide (or simply activates the enzyme towards the substrate). The peptide includes a fluorogenic donor radical, e.g., a fluorescence emitting radical, and an acceptor radical, e.g., an aromatic radical which absorbs the fluorescence energy of the fluorogenic donor radical when the acceptor radical and the fluorogenic donor radical are covalently held in close proximity. See, for example, USSN 5,527,681; 5,506,115, 5,429,766, 5,424,186, and 5,316,691; and Capobianco et al. (1992) *Anal Biochem* 204:96-102. For example, the substrate

modified endogenous gene, or a part of a completely heterologous construct, e.g., as part of a reporter gene construct.

In one embodiment, the indicator gene is an unmodified endogenous gene. In certain instances, it may be desirable to increase the level of transcriptional activation of the endogenous indicator gene by the signal pathway in order to, for example, improve the signal-to-noise of the test system, or to adjust the level of response to a level suitable for a particular detection technique. In one embodiment, the transcriptional activation ability of the signal pathway can be amplified by the overexpression of one or more of the proteins involved in the intracellular signal cascade, particularly enzymes involved in the pathway. For example, increased expression of Jun kinases (JNKs) can potentiate the level of transcriptional activation by a signal in an MEK/MEKK pathway. This approach can, of course, also be used to potentiate the level of transcription of a heterologous reporter gene as well.

In other embodiments, the sensitivity of an endogenous indicator gene can be enhanced by manipulating the promoter sequence at the natural locus for the indicator gene. Such manipulation may range from point mutations to the endogenous regulatory elements to gross replacement of all or substantial portions of the regulatory elements. In general, manipulation of the genomic sequence for the indicator gene can be carried out using techniques known in the art, including homologous recombination.

In another exemplary embodiment, the promoter (or other transcriptional regulatory sequences) of the endogenous gene can be "switched out" with a heterologous promoter sequence, e.g., to form a chimeric gene at the indicator gene locus. Again, using such techniques as homologous recombination, the regulatory sequence can be so altered at the genomic locus of the indicator gene.

In still another embodiment, a heterologous reporter gene construct can be used to provide the function of an indicator gene. Reporter gene constructs are prepared by operatively linking a reporter gene with at least one transcriptional regulatory element. If only one transcriptional regulatory element is included it must be a regulatable promoter, At least one the selected transcriptional regulatory elements must be indirectly or directly regulated by the activity of the selected cell-surface receptor whereby activity of the receptor can be monitored via transcription of the reporter genes.





Other promoters and transcriptional control elements, in addition to those described above, include the vasoactive intestinal peptide (VIP) gene promoter (cAMP responsive; Fink et al. (1988), Proc. Natl. Acad. Sci. 85:6662-6666); the somatostatin gene promoter (cAMP responsive; Montminy et al. (1986), Proc. Natl. Acad. Sci. 8.3:6682-6686); the proenkephalin promoter (responsive to cAMP, nicotinic agonists, and phorbol esters; Comb et al. (1986), Nature 323:353-356); the phosphoenolpyruvate carboxy-kinase gene promoter (cAMP responsive; Short et al. (1986), J. Biol. Chem. 261:9721-9726); the NGFI-A gene promoter (responsive to NGF, cAMP, and serum; Changelian et al. (1989). Proc. Natl. Acad. Sci. 86:377-381); and others that may be known to or prepared by those of skill in the art.

In the case of receptors which modulate cyclic AMP, a transcriptional based readout can be constructed using the cyclic AMP response element binding protein, CREB, which is a transcription factor whose activity is regulated by phosphorylation at a particular serine (S133). When this serine residue is phosphorylated, CREB binds to a recognition sequence known as a CRE (cAMP Responsive Element) found to the 5' of promotors known to be responsive to elevated cAMP levels. Upon binding of phosphorylated CREB to a CRE, transcription from this promoter is increased. Phosphorylation of CREB is seen in response to both increased cAMP levels and increased intracellular Ca levels. Increased cAMP levels result in activation of PKA, which in turn phosphorylates CREB and leads to binding to CRE and transcriptional activation. Increased intracellular calcium levels results in activation of calcium/calmodulin responsive kinase IV (CaM kinase IV). Phosphorylation of CREB by CaM kinase IV is effectively the same as phosphorylation of CREB by PKA, and results in transcriptional activation of CRE containing promotors. Activation of extracellular signal-related protein kinase (ERK) and Rsk2 by also leads to the phosphorylation and transactivation of CREB. Impey et al. (1998) Neuron 21:869-883.

Therefore, a transcriptional-based readout can be constructed in cells containing a reporter gene whose expression is driven by a basal promoter containing one or more CRE. Changes in the intracellular concentration of $Ca^{++}$ (a result of alterations in the activity of the receptor upon engagement with a ligand) will result in changes in the level of expression of the reporter gene if: a) CREB is also co-expressed in the cell, and b) either the endogenous yeast CaM kinase will phosphorylate CREB in response to

term "receptor-responsive promoter" indicates a promoter which is regulated by some product of the target receptor's signal transduction pathway.

Alternatively, the promoter may be one which is repressed by the receptor pathway, thereby preventing expression of a product which is deleterious to the cell. With a receptor repressed promoter, one screens for agonists by linking the promoter to a deleterious gene, and for antagonists, by linking it to a beneficial gene. Repression may be achieved by operably linking a receptor- induced promoter to a gene encoding mRNA which is antisense to at least a portion of the mRNA encoded by the marker gene (whether in the coding or flanking regions), so as to inhibit translation of that mRNA. Repression may also be obtained by linking a receptor-induced promoter to a gene encoding a DNA binding repressor protein, and incorporating a suitable operator site into the promoter or other suitable region of the marker gene.

h. Exemplary emodiments: Host Cells

Suitable host cells for generating the subject assay include prokaryotes, yeast, or higher eukaryotic cells, especially mammalian cells. Prokaryotes include gram negative or gram positive organisms. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651) (Gluzman (1981) Cell 23:175) CV-1 cells (ATCC CCL 70), L cells, C127, 3T3, Chinese hamster ovary (CHO), HeLa and BHK cell lines. It will be understood that to achieve selection or screening, the host cell must have an appropriate phenotype.

If yeast cells are used, the yeast may be of any species which are cultivable and in which an exogenous receptor can be made to engage the appropriate signal transduction machinery of the host cell. Suitable species include *Kluyveri lactis, Schizosaccharomyces pombe*, and *Ustilaqo maydis; Saccharomyces cerevistae* is preferred. Other yeast which can be used in practicing the present invention are *Neurospora crassa, Aspergillus niger, Aspergillus nidulans, Pichia pastoris, Candida tropicalis, and Hansenula polymorpha.* The term "yeast", as used herein, includes not only yeast in a strictly taxonomic sense, i.e., unicellular organisms, but also yeast-like multicellular fungi or filamentous fungi.

The choice of appropriate host cell will also be influenced by the choice of detection signal. For instance, reporter constructs, as described below, can provide a selectable or screenable trait upon transcriptional activation (or inactivation) in response





In the methods of the present invention, the lesion mammal can have a lesion of the fornix or a related brain structure that disrupts memory consolidation (e.g., perirhinal cortex, amygdala, medial septal nucleus, locus coeruleus, hippocampus, mammallary bodies). Lesions in the mammal can be produced by mechanical or chemical disruption. For example, the fornix lesion can be caused by surgical ablation, electrolytic, neurotoxic and other chemical ablation techniques, or reversible inactivation such as by injection of an anesthetic, e.g., tetrodooxin or lidocaine, to temporarily arrest activity in the fornix.

To further illustrate, fimbrio-fornix (rodents) and fornix (primates) lesions can be created by stereotatic ablation. In particular, neurons of the fornix structure are axotomized, e.g., by transection or aspiration (suction) ablation. A complete t[redacted]n of the fornix disrupts cholinergic and GABAergic function and electrical activity, and induces morphological reorganization in the hippocampal formation. In general, the fornix transection utilized in the subject method will not disconnect the parahippocampal region from the neocortex. In those embodiments, the fornix transection will not disrupt functions that can be carried out by the parahippocampal region independent of processing by the hippocampal formation, and hence would not be expected to produce the full-blown amnesia seen following more complete hippocampal system damage.

In one embodiment, the animal can be a rat. Briefly, the animals are anesthetized, e.g., with intraperitoneal injections of a ketamine-xylazine mixture and positioned in a Kopf stereotoxic instrument. A sagittal incision is made in the scalp and a craniotomy is performed extending 2.0 mm posterior and 3.0 mm lateral from Bregma. An aspirative device, e.g., with a 20 gauge tip, is mounted to a stereotaxic frame (Kopf Instruments) and fimbria-fornix is aspirated by placing the suction tip at the correct sterotaxic location in the animals brain. Unilateral aspirative lesions are made by suction through the cingulate cortex, completely transecting the fimbria fornix unilaterally, and (optionally) removing the dorsal tip of the hippocampus as well as the overlying cingulate cortex to inflict a partial denervation on the hippocampus target. See al[redacted] Gage et al., (1983) Brain Res. 268:27 and Gage et al. (1986) Neuroscience 19:24[redacted]

In another exemplary embodiment, the animal can be a monkey. The animals can be anesthetized, e.g., with isoflurane (1.5-2.0%). Following pretreatment with mannitol (0.25 g/kg, iv), unilateral transections of the left fornix an be performed, such

animal is placed in the lit chamber for some period of time, then the door is opened. the animal moves to the dark chamber after a short delay-the latency, that is recorded. Upon entry into the dark chamber, the door is shut closed and a footshock is delivered. Retention of the experience is determined after various time intervals, e.g., 24 or 48 hours, by repeating the test and recording the latency. The protocol is one of many variants of the inhibitory avoidance procedures (for review, see Rush (1988) Behav Neural Biol 50:255).

An exemplary maze testing embodiment is the water maze working memory test. In general, the method utilizes an apparatus which consists of a circular water tank. The water in the tank is made cloudy by the addition of milk powder. A clear plexiglass platform, supported by a movable stand rest on the bottom of the tank, is submerged just below the water surface. Normally a swimming rat cannot perceive the location of the platform but it may recall it from a previous experience and training, unless it suffers from some memory impairment. The time taken to locate the platform is measured and referred to as the latency. During the experiment, all orientational cues such as ceiling lights etc. remain unchanged. Longer latencies are generally observed with rats with some impairment to their memory.

Another memory test includes the eyeblink conditioning test, which involves the administration of white noise or steady tone that preceeds a mild air puff which stimulates the subject's eyeblink.

Still another memory test which can be used is fear conditioning, e.g., either "cued" and "contextual" fear conditioning. In one embodiment, a freeze monitor administers a sequence of stimuli (sounds, shock) and then records a series of latencies measuring the recovery from shock induced freezing of the animal.

Another memory test for the lesioned animals is a holeboard test, which utilizes a rotating holeboard apparatus containing (four) open holes arranged in a 4-corner configuration in the floor of the test enclosure. A mouse is trained to poke its head into a hole and retrieve a food reward from a "baited" hole which contains a reward on every trial. There is a food reward (e.g. Froot Loop) in every exposed hole which is made inaccessible by being placed under a screen. The screen allows the odor of the reward to emanate from the hole, but does not allow access to the reinforcer. When an individual hole is baited, a small piece of Froot Loop is placed on top of the screen, where it is





such as, but not exclusively, mannitol or glycine may be used and appropriate buffered solutions of the desired volume will be provided so as to obtain adequate isotonic buffered solutions of the desired pH. Similar solutions may also be used for the pharmaceutical compositions of compounds in isotonic solutions of the desired volume and include, but not exclusively, the use of buffered saline solutions with phosphate or citrate at suitable concentrations so as to obtain at all times isotonic pharmaceutical preparations of the desired pH, (for example, neutral pH).

In certain embodiment, the pharmaceutical of the present invention is a gene delivery system for gene therapy with a therapeutic LTM gene. Such gene therapy systems can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Pat. No. *5,328,470*) or by stereotactic injection (e.g. Chen et al. (1994) Proc. Natl. Acad. Sci. USA 91: 3054-3057).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

*F. Methods of Treatment*

In various embodiments, the present invention contemplates modes of treatment and prophylaxis which utilitize one or more of the subject LTM genes (e.g., by gene therapy) or antisense constructs thereto, the LTM proteins (e.g., for protein therapy) or peptidomimetics thereof, or compounds identified in the subject drug screening assays. These agents may be useful for altering (increasing or decreasing) the occurrence of

used to treat amensias of longer duration, such as post concussive or as the result of Herpes simplex encephalitis.

*(i) Effective Dose*

Toxicity and therapeutic efficacy of compounds to be used in the treatment methods of the present invention can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining The LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

*G. Diagnostic and Prognostic Assays*

The present method also provides a method for determining if a subject is at risk for a disorder characterized deterioration of memory consolidation. In preferred embodiments, the methods can be characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding an LTM protein, or (ii) the mis-expression of the LTM gene. To illustrate, such genetic lesions can be





In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Pat. Nos. *4,683,195* and *4,683,202)*, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) PNAS 91:360-364), the latter of which can be particularly useful for detecting point mutations in the LTM gene (see Abravaya et al. (1995) Nuc Acid Res 23:675-682). In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to an LTM gene under conditions such that hybridization and amplification of the LTM gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Another embodiment of the invention provides for a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of an LTM gene, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject LTM genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels. Such oligonucleotide probes can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, deterioration in memeory consolidation.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving memory or an LTM gene.

material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One means for labeling an anti-LTM protein specific antibody is via linkage to an enzyme and use in an enzyme immunoassay (EIA) (Voller, "The Enzyme Linked Immunosorbent Assay (ELISA)", Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, Md.; Voller, et al., J. Clin. Pathol. 31:507-520 (1978); Butler, Meth. Enzymol. 73:482-523 (1981); Maggio, (ed.) Enzyme Immunoassay, CRC Press, Boca Raton, Fla., 1980; Ishikawa, et al., (eds.) Enzyme Immunoassay, Kgaku Shoin, Tokyo, 1981). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type or mutant peptides through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The





identify drugs, pharmaceuticals, therapies and interventions which may be effective in treating disease.

One aspect of the present invention concerns the use of a transgenic animal which is comprised of cells (of that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous LTM protein in one or more cells in the animal. An LTM transgene can encode the wild-type form of the protein, or can encode homologs thereof including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of an LTM protein can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, lack of LTM expression which might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this and, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BL/6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, Me.). Preferred strains are those with H-$2^b$, H-$2^d$ or H-$2^q$ haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice this invention may themselves be transgenics, and/or may be knockouts (i.e., obtained from animals which have one or more genes partially or (completely suppressed).

(Alberini, 1994). Based on these data, we proposed that memory-inducing stimuli activate a gene cascade, where CREB regulates the expression of regulatory IEGs which, in turn, regulate the expression of more downstream target genes required for long-term memory (Alberini, 1994). This model accounts for why gene expression seems to be required only for an early and brief time window. The essential gene expression appears to be brief because it corresponds to the critical time necessary for the expression of regulatory IEGs, which would be the rate-limiting step of the molecular cascade of events leading to long-term memory.

In mammals, changes in mRNA and protein levels of IEGs such as c-fos and zif268 have been widely investigated in memory and models of synaptic plasticity. Although still controversial (Campeau, 1991), particularly due to the lack of precise controls or quantitative measurements (the studies have been generally based on immunohistochemical analysis), both c-fos and zif268 have been reported to be up-regulated in several brain areas following different kinds of learning (rev. in Dragunow, 1996). Therefore, in our initial pilot experiments, we performed a Northern blot analysis of changes in hippocampal zif268 and c-fos mRNA levels following IA training, focusing on the temporal window suggested by induction of PCREB; i.e., 3, 6, 9, and 20 hrs after training. Three animals per timepoint were investigated, and the hybridizations were normalized using cyclophilin gene as a control probe. The same membrane was sequentially hybridized with all the probes. As shown in Fig. 1, we found that zif268 was induced in all the trained animals at 9 and 20 hr after training. On the contrary, c-fos did not show any evident change throughout the timecourse. Thus, the IEG response to IA training are selective with respect to the genes activated as well as the time following training.

Fig. 1. Time-course Northern blot analysis of zif268 and c-fos following IA training. Increase in zif268, but not c-fos mRNA is evident in all animals at 9 and 20 hr after training. These results showed that zif268 is induced in the hippocampus by IA. They also reveal *when* the induction of an IEG is detectable in the hippocampus following IA, namely at 9 and 20 hrs. On the basis of these data, as described below, we decided to perform our first array hybridization analysis using the RNA obtained from animals 9 hr after training.





**Example 2**

**Array hybridization reveals new gene responses to IA training.**

We believe that our model system lends itself particularly well to searching for genes differentially expressed in long-term memory. We have determined where and when to carry out the differential analysis in animals that have the ability to learn vs. those with memory impairment. We also know that learning induces a hippocampal gene response that is sensitive to fornix lesions.

The possibility of identifying differentially expressed genes is one of the most powerful approaches for understanding the gene pattern linked to a specific function. A multitude of techniques have become available in recent years to isolate differentially expressed genes and some have been successfully used in isolating genes involved in long-term memory (Cavallaro, 1997). The most advanced generation of these techniques is the hybridization of DNA arrays. The arrays consist of supports (nylon or glass) on which cDNA fragments have been immobilized systematically. A schematic representation of how this technique works is shown in Figure 3.

Figure 3. Broad scale expression profiling with cDNA expression arrays. Side by side hybridizations with cDNA probes prepared from two different RNA populations allow the simultaneous comparison of the expression levels of all the cDNAs on the array. (from Clontech user manual). Arrays can contain, in principle, all the cloned DNA sequences. We begun our analysis using a relatively small array, the AtlasTM rat cDNA expression array from Clontech (PaloAlto, Ca), which was the most complete commercially available array at that time. This array contained 588 genes isolated from rat brain and coding for a variety of molecules involved in the pathways that regulate brain function, such as signaling molecules, receptors, signal transduction proteins, extracellular proteins, structural molecules, molecules involved in synaptic transmission, and molecules involved in neural pathologies, including Alzheimer disease. The DNAs were fixed on a positively charged nylon membrane.

Since it is possible to detect changes in mRNA expression at 9 hr after IA training compared to controls, we began the search for differential gene expression memory using array hybridization and compared animals that walked through the apparatus without receiving a shock (0 no shock) and were immediately sacrificed to animals that underwent IA training and were sacrificed 9 hr later. Hippocampi from 4

These results show that the regulation of genes #2 and #3 is selectively associated with IA training. As with C/EBPβ, the regulation of gene #2 that occurred with learning was abolished in fornix-lesioned animals, although the basal expression of the gene was not affected by lesioning. These data confirmed the IA memory formation is accompanied by regulation of gene expression that is induced by inputs passing through the fornix. Gene #3, on the other hand, appeared to be down-regulated with IA memory as well as in all animals with fornix lesions. From these data we conclude that fornix lesions produce changes in the regulation of certain genes within the hippocampus. This result is intriguing and suggests that the fornix may contribute to memory formation by modulating either constitutive or IA-induced gene expression.

Taken together, these data demonstrate that we are able to detect and analyze changes in gene expression after IA training. We have confirmed that some of these changes do not occur in the hippocampi of animals with lesions of the fornix which have impaired memory consolidation. We believe that these preliminary data show convincingly that our IA model is suitable for the analysis of gene expression changes in long-term memory.

We recently began to establish hybridization conditions of more complete arrays that have very recently become available from Research Genetics. These arrays contain 5,000 rat transcripts. Our goal is to carry out a systematic analysis of the rat cloned genes and identify which change their expression following training. To screen for genes whose expression change at different times after learning, as described in the Research Plan section, we plan to analyze several time windows after training and follow up with Northern blot analyses. Moreover, to define whether the gene cascade activated in IA is a general molecular mechanism of memory, we will analyze the expression of the identified genes in other forms of memory, including contextual fear conditioning and Morris water maze.

**Example 3**

**The delayed and prolonged gene response during memory consolidation makes feasible detection of changes after the Morris water maze.**

One hopes that the genes involved in long-term memory consolidation are conserved, not only evolutionarily, but also across different types of memory that utilize the same neural structures. We chose IA training originally because it could be used to





Figure 6. Panel A. Mean escape latency for rats trained on the water maze. Rats received eight trials a day, for four consecutive days. Panel B. Mean escape latency on each of the eight trials of day one (trials 1-8) and day two (trials 9-16).

**Example 4**

**Changes in gene expression detected by hybridizations of array.**

In this project we will identify which genes, among approximately 5,000 transcripts including genes and expressed sequence tag (ESTs), are regulated in the hippocampus during IA memory formation. We will compare parallel hybridizations of identical arrays with cDNA probes obtained from hippocampal RNA of untrained and trained animals. The transcripts that will show a significant change in expression in the trained condition ws the untrained will be further analyzed by Northern blot, following the sequence of steps described above in 1 to 4.

The array hybridization screening for genes differentially expressed following IA training will be carried on arrays purchased from Research Genetics (Huntsville, AL). This company has recently released the most complete rat DNA array commercially available, the *rat GeneFilters microarrays*. These arrays contain over 5,000 spots, representing approximately 1,700 named rat genes as well as many rat ESTs that are considered similar to named genes in other organisms. Each spot on the membrane contains approximately 0.5 ng of insert DNA from a cDNA clone containing the 3' end of a gene. The insert cDNA has been denatured and UV-cross-linked to the positively charged membrane. The manufacturer provides a detailed protocol that insures that the hybridizations are carried out under conditions where most probes (cDNA from the experimental mRNAs) are not saturating the spots. However, as described below we will set up several hybridization conditions that will provide the highest probability of success for identifying genes regulated in memory. These arrays (nylon membranes) are similar to those we purchased from Clontech and successfully used to identify several genes, as described in Preliminary Studies. However, they contain a much greater number of transcripts. Finally, Research Genetics also provides a software analysis tool that, when used in conjunction with Gene Filters, allows for the comparison of gene expression from images produced on a phosphor imaging system. This system allows for normalization across multiple experiments and has a built-in database to facilitate archiving of both raw images and fitted data. To maximize the probability of detecting

We Claim:

1. A method for modulating long term memory consolidation in an animal comprising treating an animal with an agent that modulates the activity of one or more of zif 268, insulin-like growth factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF.

2. A method for enhancing long term memory consolidation in an animal comprising treating an animal with an agent that modulates a signal transduction pathway of glutamate receptor 1 (GluR1) or glutamate receptor 2 (GluR2), which agent is a ligand for the GluR1 or GluR2 receptor.

3. A method for identifying an agent which modulates memory consolidation, comprising,
   (i) providing a reaction system for detecting the activity of a product encoded by a gene selected from the group consisting of of zif 268, insulin-like growth factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF;
   (ii) contacting said system with a test compound; and
   (iii) determining if the test compound alters the activity of the gene product.

4. A method for identifying an agent which modulates memory consolidation, comprising,
   (i) providing a reaction system for detecting the level of expression of a gene selected from the group consisting of of zif 268, insulin-like growth factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF;
   (ii) contacting said system with a test compound; and
   (iii) determining if the test compound alters the level of expression of the gene.

5. The method of claim 3 or 4, wherein the reaction system is a cell-free system.





16. The method of claim 14 or 15, further comprising administering, conjointly with the pharmaceutical preparation, one or more of a neuronal growth factor, a neuronal survival factor, and a neuronal tropic factor.

17. The method of claim 14 or 15, further comprising administering, conjointly with the pharmaceutical preparation, an agent that activates CREB-dependent transcription in an amount sufficient to produce a memory enhancing effect.

18. The method of claim 17, wherein the CREB activating agent is a cAMP elevating agent.

19. The method of claim 18, wherein at least one cAMP agonist activates a [redacted]te cyclase.

20. The method of claim 17, wherein the CREB activating agent is a cAMP analog.

21. The method of claim 17, wherein the CREB activating agent is a cAMP phosphodiesterase inhibitor.

22. A method for assessing a patient for learning and/or memory functional performance including a step of detecting the expression of a gene selected from the group consisting of of zif 268, insulin-like growth factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/EBPβ and VGF, or the level of activity of a gene product thereof, in the patient's hippocampus.

23. A method for assessing a patient for learning and/or memory functional performance including a step of detecting the expression of, or a mutation in, one or more genes selected from the group consisting of zif 268, insulin-like [redacted]th factor, glutamate receptor 1 (GluR1), glutamate receptor 2 (GluR2), c/E[redacted]d VGF, or the level of activity of the gene products thereof, (optionally) in the patient's hippocampus.

Figure 1

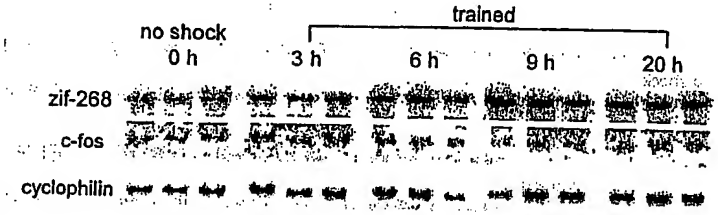






Figure 3

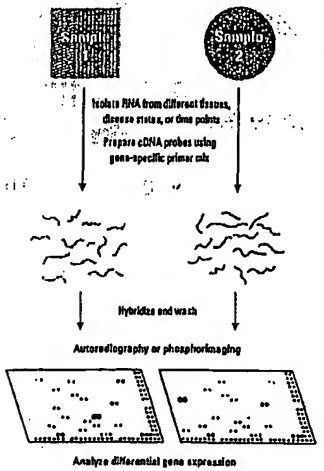

Figure 5

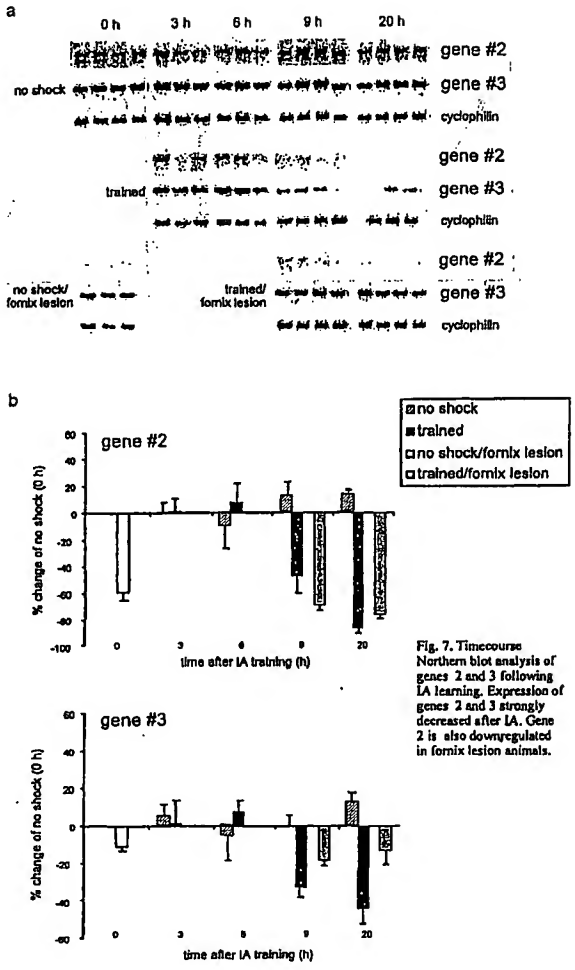


Fig. 7. Timecourse Northern blot analysis of genes 2 and 3 following IA learning. Expression of genes 2 and 3 strongly decreased after IA. Gene 2 is also downregulated in fornix lesion animals.





SEQUENCE LISTING

<110> BEAR, Mark
ALBERINI, Cristina

<120> METHODS AND COMPOSITIONS FOR REGULATING MEMORY CONSOLIDATION

<130> NIIN-pWO-008

<140> PCT/US01/000000
<141> 2000-04-02

<150> US 60/193614
<151> 2000-03-31

<160> 12

<170> PatentIn version 3.0

<210> 1
<211> 2929
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (104)..(2824)

<400> 1
cgaaaagaac aggcagaaca gcgagaagaa taaagggaaa ggggggggaaa caccaaatct
60

atgattggac ctgggcttct ttttcgccaa tccaaaaagg aat atg cag cac att
115
Met Gln His Ile
1

ttt gcc ttc ttc tgc acc ggt ttc cta ggc gcg gta gta ggt gcc aat
163
Phe Ala Phe Phe Cys Thr Gly Phe Leu Gly Ala Val Val Gly Ala Asn
5 10 15 20

ttc ccc aac aat atc cag atc ggg gga tta ttt cca aac cag cag tca
211
Phe Pro Asn Asn Ile Gln Ile Gly Gly Leu Phe Pro Asn Gln Gln Ser
25 30 35

cag gaa cat gct gct ttt aga ttt gct ttg tcg caa ctc aca gag ccc
259
Gln Glu His Ala Ala Phe Arg Phe Ala Leu Ser Gln Leu Thr Glu Pro
40 45 50

ccg aag ctg ctc ccc cag att gat att gtg aac atc agc gac agc ttt
307
Pro Lys Leu Leu Pro Gln Ile Asp Ile Val Asn Ile Ser Asp Ser Phe
55 60 65

gag atg acc tat aga ttc tgt tcc cag ttc tcc aaa gga gtc tat gcc
355
Glu Met Thr Tyr Arg Phe Cys Ser Gln Phe Ser Lys Gly Val Tyr Ala
70 75 80

atc ttt ggg ttt tat gaa cgt agg act gtc aac atg ctg acc tcc ttt
403
Ile Phe Gly Phe Tyr Glu Arg Arg Thr Val Asn Met Leu Thr Ser Phe
85 90 95 100

```
agg cag aga att gat ata tct cgc cgg ggg aat gct ggg gat tgt ctg
1075
Arg Gln Arg Ile Asp Ile Ser Arg Arg Gly Asn Ala Gly Asp Cys Leu
    310                 315                 320

gct aac cca gct gtt ccc tgg ggc caa ggg atc gac atc cag aga gct
1123
Ala Asn Pro Ala Val Pro Trp Gly Gln Gly Ile Asp Ile Gln Arg Ala
325                 330                 335                 340

ctg cag cag gtc gcg ttt gaa ggt tta aca gga aac gtg cag ttt aat
1171
Leu Gln Gln Val Ala Phe Glu Gly Leu Thr Gly Asn Val Gln Phe Asn
                345                 350                 355

gag aaa gga cgc cgg acc aac tac acg ctc cac gtg att gaa atg aaa
1219
Glu Lys Gly Arg Arg Thr Asn Tyr Thr Leu His Val Ile Glu Met Lys
            360                 365                 370

cat gac ggc atc cga aag att ggt tac tgg aat gaa gat gat aag ttt
1267
His Asp Gly Ile Arg Lys Ile Gly Tyr Trp Asn Glu Asp Asp Lys Phe
        375                 380                 385

gtc cct gca gcc acc gat gcc caa gct ggg ggg gat aat tca agt gtt
1315
Val Pro Ala Ala Thr Asp Ala Gln Ala Gly Gly Asp Asn Ser Ser Val
    390                 395                 400

cag aac aga aca tac atc gtc aca aca atc cta gaa gat cct tat gtg
1363
Gln Asn Arg Thr Tyr Ile Val Thr Thr Ile Leu Glu Asp Pro Tyr Val
405                 410                 415                 420

atg ctc aag aag aac gcc aat cag ttt gag ggc aat gac cgt tac gag
1411
Met Leu Lys Lys Asn Ala Asn Gln Phe Glu Gly Asn Asp Arg Tyr Glu
                425                 430                 435

ggc tac tgt gta gag ctg gcg gca gag att gcc aag cac gtg ggc tac
1459
Gly Tyr Cys Val Glu Leu Ala Ala Glu Ile Ala Lys His Val Gly Tyr
            440                 445                 450

tcc tac cgt ctg gag att gtc agt gat gga aaa tac gga gcc cga gac
1507
Ser Tyr Arg Leu Glu Ile Val Ser Asp Gly Lys Tyr Gly Ala Arg Asp
        455                 460                 465

cct gac acg aag gcc tgg aat ggc atg gtg gga gag ctg gtc tat gga
1555
Pro Asp Thr Lys Ala Trp Asn Gly Met Val Gly Glu Leu Val Tyr Gly
    470                 475                 480

aga gca gat gtg gct gtg gct ccc ctt act atc act ttg gtc cgg gaa
1603
Arg Ala Asp Val Ala Val Ala Pro Leu Thr Ile Thr Leu Val Arg Glu
485                 490                 495                 500

gaa gtt ata gat ttc tcc aaa cca ttt atg agt ttg ggg atc tcc atc
1651
Glu Val Ile Asp Phe Ser Lys Pro Phe Met Ser Leu Gly Ile Ser Ile
                505                 510                 515

atg att aaa aaa cca cag aaa tcc aag ccg ggt gtc ttc tcc ttc ctt
```





```
Tyr Ile Glu Gln Arg Lys Pro Cys Asp Thr Met Lys Val Gly Gly Asn
725                 730                 735                 740

ttg gat tcc aaa ggc tat ggc att gca aca ccc aag ggg tct gcc ctg
2371
Leu Asp Ser Lys Gly Tyr Gly Ile Ala Thr Pro Lys Gly Ser Ala Leu
            745                 750                 755

aga aat cca gta aac ctg gca gtg tta aaa ctg aac gag cag ggg ctt
2419
Arg Asn Pro Val Asn Leu Ala Val Leu Lys Leu Asn Glu Gln Gly Leu
        760                 765                 770

ttg gac aaa ttg aaa aac aaa tgg tgg tac gac aag ggc gag tgc ggc
2467
Leu Asp Lys Leu Lys Asn Lys Trp Trp Tyr Asp Lys Gly Glu Cys Gly
    775                 780                 785

agc ggg gga ggt gat tcc aag gac aag aca agc gct ctg agc ctc agc
2515
Ser Gly Gly Gly Asp Ser Lys Asp Lys Thr Ser Ala Leu Ser Leu Ser
790                 795                 800

aat gtg gca ggc gtg ttc tac atc ctg atc gga gga ctt gga cta gcc
2563
Asn Val Ala Gly Val Phe Tyr Ile Leu Ile Gly Gly Leu Gly Leu Ala
805                 810                 815                 820

atg ctg gtt gcc tta atc gag ttc tgc tac aaa tcc cgt agt gaa tcc
2611
Met Leu Val Ala Leu Ile Glu Phe Cys Tyr Lys Ser Arg Ser Glu Ser
                825                 830                 835

aag cgg atg aag ggt ttt tgt ttg atc cca cag caa tcc atc aac gaa
2659
Lys Arg Met Lys Gly Phe Cys Leu Ile Pro Gln Gln Ser Ile Asn Glu
            840                 845                 850

gcc ata cgg aca tcg acc ctc ccc cgc aac agc ggg gca gga gcc agc
2707
Ala Ile Arg Thr Ser Thr Leu Pro Arg Asn Ser Gly Ala Gly Ala Ser
        855                 860                 865

agc ggc ggc agt gga gag aat ggt cgg gtg gtc agc cat gac ttc ccc
2755
Ser Gly Gly Ser Gly Glu Asn Gly Arg Val Val Ser His Asp Phe Pro
    870                 875                 880

aag tcc atg caa tcg att cct tgc atg agc cac agt tca ggg atg ccc
2803
Lys Ser Met Gln Ser Ile Pro Cys Met Ser His Ser Ser Gly Met Pro
885                 890                 895                 900

ttg gga gcc acg gga ttg taa ctggagcaga tggagacccc ttggggagca
2854
Leu Gly Ala Thr Gly Leu
            905

ggctcgggct ccccagcccc atcccaaacc cttcagtgcc aaaaacaaca acaaaatgaa
2914

acgcaaccgg aattc
2929


<210> 2
```

```
Asn Lys Phe Lys Glu Ser Gly Ala Asn Val Thr Gly Phe Gln Leu Val
                245                 250                 255

Asn Tyr Thr Asp Thr Ile Pro Ala Lys Ile Met Gln Gln Trp Lys Asn
            260                 265                 270

Ser Asp Ala Arg Asp His Thr Arg Val Asp Trp Lys Arg Pro Lys Tyr
        275                 280                 285

Thr Ser Ala Leu Thr Tyr Asp Gly Val Lys Val Met Ala Glu Ala Phe
    290                 295                 300

Gln Ser Leu Arg Arg Gln Arg Ile Asp Ile Ser Arg Arg Gly Asn Ala
305                 310                 315                 320

Gly Asp Cys Leu Ala Asn Pro Ala Val Pro Trp Gly Gln Gly Ile Asp
                325                 330                 335

Ile Gln Arg Ala Leu Gln Gln Val Ala Phe Glu Gly Leu Thr Gly Asn
            340                 345                 350

Val Gln Phe Asn Glu Lys Gly Arg Arg Thr Asn Tyr Thr Leu His Val
        355                 360                 365

Ile Glu Met Lys His Asp Gly Ile Arg Lys Ile Gly Tyr Trp Asn Glu
    370                 375                 380

Asp Asp Lys Phe Val Pro Ala Ala Thr Asp Ala Gln Ala Gly Gly Asp
385                 390                 395                 400

Asn Ser Ser Val Gln Asn Arg Thr Tyr Ile Val Thr Thr Ile Leu Glu
                405                 410                 415

Asp Pro Tyr Val Met Leu Lys Lys Asn Ala Asn Gln Phe Glu Gly Asn
            420                 425                 430

Asp Arg Tyr Glu Gly Tyr Cys Val Glu Leu Ala Ala Glu Ile Ala Lys
        435                 440                 445

His Val Gly Tyr Ser Tyr Arg Leu Glu Ile Val Ser Asp Gly Lys Tyr
    450                 455                 460

Gly Ala Arg Asp Pro Asp Thr Lys Ala Trp Asn Gly Met Val Gly Glu
465                 470                 475                 480

Leu Val Tyr Gly Arg Ala Asp Val Ala Val Ala Pro Leu Thr Ile Thr
                485                 490                 495

Leu Val Arg Glu Glu Val Ile Asp Phe Ser Lys Pro Phe Met Ser Leu
            500                 505                 510
```





```
Glu Gln Gly Leu Leu Asp Lys Leu Lys Asn Lys Trp Trp Tyr Asp Lys
    770                 775                 780

Gly Glu Cys Gly Ser Gly Gly Gly Asp Ser Lys Asp Lys Thr Ser Ala
785                 790                 795                 800

Leu Ser Leu Ser Asn Val Ala Gly Val Phe Tyr Ile Leu Ile Gly Gly
                805                 810                 815

Leu Gly Leu Ala Met Leu Val Ala Leu Ile Glu Phe Cys Tyr Lys Ser
            820                 825                 830

Arg Ser Glu Ser Lys Arg Met Lys Gly Phe Cys Leu Ile Pro Gln Gln
        835                 840                 845

Ser Ile Asn Glu Ala Ile Arg Thr Ser Thr Leu Pro Arg Asn Ser Gly
    850                 855                 860

Ala Gly Ala Ser Ser Gly Gly Ser Gly Glu Asn Gly Arg Val Val Ser
865                 870                 875                 880

His Asp Phe Pro Lys Ser Met Gln Ser Ile Pro Cys Met Ser His Ser
                885                 890                 895

Ser Gly Met Pro Leu Gly Ala Thr Gly Leu
            900                 905


<210>  3
<211>  2621
<212>  DNA
<213>  Homo sapiens

<220>
<221>  CDS
<222>  (3)..(2621)

<400>  3
cc atg gga tcg ctg ctt gcg ctc ctg gca ctg ctg ccg ctg tgg ggt
47
   Met Gly Ser Leu Leu Ala Leu Leu Ala Leu Leu Pro Leu Trp Gly
   1               5                  10                  15

gct gtg gct gag ggc cca gcc aag aag gtg ctg acc ctg gag gga gac
95
Ala Val Ala Glu Gly Pro Ala Lys Lys Val Leu Thr Leu Glu Gly Asp
                20                  25                  30

ttg gtg ctg ggt ggg ctg ttc cca gtg cac cag aag ggc ggc cca gca
143
Leu Val Leu Gly Gly Leu Phe Pro Val His Gln Lys Gly Gly Pro Ala
            35                  40                  45

gag gac tgt ggt cct gtc aat gag cac cgt ggc atc cag cgc ctg gag
191
Glu Asp Cys Gly Pro Val Asn Glu His Arg Gly Ile Gln Arg Leu Glu
        50                  55                  60
```

```
863
Ser Glu Asp Ala Arg Glu Leu Leu Ala Ala Ser Gln Arg Leu Asn Ala
                275                 280                 285

agc ttc acc tgg gtg gcc agt gat ggt tgg ggg gcc ctg gag agt gtg
911
Ser Phe Thr Trp Val Ala Ser Asp Gly Trp Gly Ala Leu Glu Ser Val
        290                 295                 300

gtg gca ggc agt gag ggg gct gct gag ggt gct atc acc atc gag ctg
959
Val Ala Gly Ser Glu Gly Ala Ala Glu Gly Ala Ile Thr Ile Glu Leu
    305                 310                 315

gcc tcc tac ccc atc agt gac ttt gcc tcc tac ttc cag agc ctg gac
1007
Ala Ser Tyr Pro Ile Ser Asp Phe Ala Ser Tyr Phe Gln Ser Leu Asp
320                 325                 330                 335

cct tgg aac aac agc cgg aac ccc tgg ttc cgt gaa ttc tgg gag cag
1055
Pro Trp Asn Asn Ser Arg Asn Pro Trp Phe Arg Glu Phe Trp Glu Gln
                340                 345                 350

agg ttc cgc tgc agc ttc cgg cag cga gac tgc gca gcc cac tct ctc
1103
Arg Phe Arg Cys Ser Phe Arg Gln Arg Asp Cys Ala Ala His Ser Leu
            355                 360                 365

cgg gct gtg ccc ttt gaa cag gag tcc aag atc atg ttt gtg gtc aat
1151
Arg Ala Val Pro Phe Glu Gln Glu Ser Lys Ile Met Phe Val Val Asn
        370                 375                 380

gca gtg tac gcc atg gcc cat gcg ctc cac aac atg cac cgt gcc ctc
1199
Ala Val Tyr Ala Met Ala His Ala Leu His Asn Met His Arg Ala Leu
    385                 390                 395

tgc ccc aac acc acc cgg ctc tgt gac gcg atg cgg cca gtt aac ggg
1247
Cys Pro Asn Thr Thr Arg Leu Cys Asp Ala Met Arg Pro Val Asn Gly
400                 405                 410                 415

cgc cgc ctc tac aag gac ttt gtg ctc aac gtc aag ttt gat gcc ccc
1295
Arg Arg Leu Tyr Lys Asp Phe Val Leu Asn Val Lys Phe Asp Ala Pro
                420                 425                 430

ttt cgc cca gct gac acc cac aat gag gtc cgc ttt gac cgc ttt ggt
1343
Phe Arg Pro Ala Asp Thr His Asn Glu Val Arg Phe Asp Arg Phe Gly
            435                 440                 445

gat ggt att ggc cgc tac aac atc ttc acc tat ctg cgt gca ggc agt
1391
Asp Gly Ile Gly Arg Tyr Asn Ile Phe Thr Tyr Leu Arg Ala Gly Ser
        450                 455                 460

ggg cgc tat cgc tac cag aag gtg ggc tac tgg gca gaa ggc ttg act
1439
Gly Arg Tyr Arg Tyr Gln Lys Val Gly Tyr Trp Ala Glu Gly Leu Thr
    465                 470                 475

ctg gac acc agc ctc atc cca tgg gcc tca ccg tca gcc ggc ccc ctg
1487
```





```
      690                 695                 700

ggc aca ggc aag gag aca gcc ccc gaa cgg cgg gag gtg gtg aca ctg
2159
Gly Thr Gly Lys Glu Thr Ala Pro Glu Arg Arg Glu Val Val Thr Leu
    705                 710                 715

cgc tgc aac cac cgc gat gca agt atg ttg ggc tcg ctg gcc tac aat
2207
Arg Cys Asn His Arg Asp Ala Ser Met Leu Gly Ser Leu Ala Tyr Asn
720                 725                 730                 735

gtg ctc ctc atc gcg ctc tgc acg ctt tat gcc ttc aat act cgc aag
2255
Val Leu Leu Ile Ala Leu Cys Thr Leu Tyr Ala Phe Asn Thr Arg Lys
                740                 745                 750

tgc ccc gaa aac ttc aac gag gcc aag ttc att ggc ttc acc atg tac
2303
Cys Pro Glu Asn Phe Asn Glu Ala Lys Phe Ile Gly Phe Thr Met Tyr
            755                 760                 765

acc acc tgc atc atc tgg ctg gca ttg ttg ccc atc ttc tat gtc acc
2351
Thr Thr Cys Ile Ile Trp Leu Ala Leu Leu Pro Ile Phe Tyr Val Thr
        770                 775                 780

tcc agt gac tac cgg gta cag acc acc acc atg tgc gtg tca gtc agc
2399
Ser Ser Asp Tyr Arg Val Gln Thr Thr Thr Met Cys Val Ser Val Ser
    785                 790                 795

ctc agc ggc tcc gtg gtg ctt ggc tgc ctc ttt gcg ccc aag ctg cac
2447
Leu Ser Gly Ser Val Val Leu Gly Cys Leu Phe Ala Pro Lys Leu His
800                 805                 810                 815

atc atc ctc ttc cag ccg cag aag aac gtg gtt agc cac cgg gca ccc
2495
Ile Ile Leu Phe Gln Pro Gln Lys Asn Val Val Ser His Arg Ala Pro
                820                 825                 830

acc agc cgc ttt ggc agt gct gct gcc agg gcc agc tcc agc ctt ggc
2543
Thr Ser Arg Phe Gly Ser Ala Ala Ala Arg Ala Ser Ser Ser Leu Gly
            835                 840                 845

caa ggg tct ggc tcc cag ttt gtc ccc act gtt tgc aat ggc cgt gag
2591
Gln Gly Ser Gly Ser Gln Phe Val Pro Thr Val Cys Asn Gly Arg Glu
        850                 855                 860

gtg gtg gac tcg aca acg tca tcg ctt tga
2621
Val Val Asp Ser Thr Thr Ser Ser Leu
    865                 870


<210>  4
<211>  872
<212>  PRT
<213>  Homo sapiens

<400>  4

Met Gly Ser Leu Leu Ala Leu Leu Ala Leu Leu Pro Leu Trp Gly Ala
```

```
Glu Asp Ala Arg Glu Leu Leu Ala Ala Ser Gln Arg Leu Asn Ala Ser
        275                 280                 285

Phe Thr Trp Val Ala Ser Asp Gly Trp Gly Ala Leu Glu Ser Val Val
    290                 295                 300

Ala Gly Ser Glu Gly Ala Ala Glu Gly Ala Ile Thr Ile Glu Leu Ala
305                 310                 315                 320

Ser Tyr Pro Ile Ser Asp Phe Ala Ser Tyr Phe Gln Ser Leu Asp Pro
                325                 330                 335

Trp Asn Asn Ser Arg Asn Pro Trp Phe Arg Glu Phe Trp Glu Gln Arg
            340                 345                 350

Phe Arg Cys Ser Phe Arg Gln Arg Asp Cys Ala Ala His Ser Leu Arg
        355                 360                 365

Ala Val Pro Phe Glu Gln Glu Ser Lys Ile Met Phe Val Val Asn Ala
    370                 375                 380

Val Tyr Ala Met Ala His Ala Leu His Asn Met His Arg Ala Leu Cys
385                 390                 395                 400

Pro Asn Thr Thr Arg Leu Cys Asp Ala Met Arg Pro Val Asn Gly Arg
                405                 410                 415

Arg Leu Tyr Lys Asp Phe Val Leu Asn Val Lys Phe Asp Ala Pro Phe
            420                 425                 430

Arg Pro Ala Asp Thr His Asn Glu Val Arg Phe Asp Arg Phe Gly Asp
        435                 440                 445

Gly Ile Gly Arg Tyr Asn Ile Phe Thr Tyr Leu Arg Ala Gly Ser Gly
    450                 455                 460

Arg Tyr Arg Tyr Gln Lys Val Gly Tyr Trp Ala Glu Gly Leu Thr Leu
465                 470                 475                 480

Asp Thr Ser Leu Ile Pro Trp Ala Ser Pro Ser Ala Gly Pro Leu Ala
                485                 490                 495

Ala Ser Arg Cys Ser Glu Pro Cys Leu Gln Asn Glu Val Lys Ser Val
            500                 505                 510

Gln Pro Gly Glu Val Cys Cys Trp Leu Cys Ile Pro Cys Gln Pro Tyr
        515                 520                 525

Glu Tyr Arg Leu Asp Glu Phe Thr Cys Ala Asp Cys Gly Leu Gly Tyr
```





```
Ser Gly Ser Val Val Leu Gly Cys Leu Phe Ala Pro Lys Leu His Ile
                805                 810                 815

Ile Leu Phe Gln Pro Gln Lys Asn Val Val Ser His Arg Ala Pro Thr
            820                 825                 830

Ser Arg Phe Gly Ser Ala Ala Ala Arg Ala Ser Ser Ser Leu Gly Gln
        835                 840                 845

Gly Ser Gly Ser Gln Phe Val Pro Thr Val Cys Asn Gly Arg Glu Val
    850                 855                 860

Val Asp Ser Thr Thr Ser Ser Leu
865                 870
```

<210> 5
<211> 3132
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (271)..(1902)

<400> 5

```
ccgcagaact tggggagccg ccgccgccat ccgccgccgc agccagcttc cgccgccgca
60

ggaccggccc ctgccccagc ctccgcagcc gcggcgcgtc cacgcccgcc cgcgcccagg
120

gcgagtcggg gtcgccgcct gcacgcttct cagtgttccc cgcgccccgc atgtaacccg
180

gccaggcccc cgcaacggtg tcccctgcag ctccagcccc gggctgcacc cccccgcccc
240

gacaccagct ctccagcctg ctcgtccagg atg gcc gcg gcc aag gcc gag atg
294
                                    Met Ala Ala Ala Lys Ala Glu Met
                                    1               5

cag ctg atg tcc ccg ctg cag atc tct gac ccg ttc gga tcc ttt cct
342
Gln Leu Met Ser Pro Leu Gln Ile Ser Asp Pro Phe Gly Ser Phe Pro
    10                  15                  20

cac tcg ccc acc atg gac aac tac cct aag ctg gag gag atg atg ctg
390
His Ser Pro Thr Met Asp Asn Tyr Pro Lys Leu Glu Glu Met Met Leu
25                  30                  35                  40

ctg agc aac ggg gct ccc cag ttc ctc ggc gcc gcc ggg gcc cca gag
438
Leu Ser Asn Gly Ala Pro Gln Phe Leu Gly Ala Ala Gly Ala Pro Glu
                45                  50                  55

ggc agc ggc agc aac agc agc agc agc agc agc ggg ggc ggt gga ggc
486
Gly Ser Gly Ser Asn Ser Ser Ser Ser Ser Ser Gly Gly Gly Gly Gly
```

```
acc cag cag cct tcg cta acc cct ctg tct act att aag gcc ttt gcc
1158
Thr Gln Gln Pro Ser Leu Thr Pro Leu Ser Thr Ile Lys Ala Phe Ala
                285                 290                 295

act cag tcg ggc tcc cag gac ctg aag gcc ctc aat acc agc tac cag
1206
Thr Gln Ser Gly Ser Gln Asp Leu Lys Ala Leu Asn Thr Ser Tyr Gln
            300                 305                 310

tcc cag ctc atc aaa ccc agc cgc atg cgc aag tat ccc aac cgg ccc
1254
Ser Gln Leu Ile Lys Pro Ser Arg Met Arg Lys Tyr Pro Asn Arg Pro
        315                 320                 325

agc aag acg ccc ccc cac gaa cgc cct tac gct tgc cca gtg gag tcc
1302
Ser Lys Thr Pro Pro His Glu Arg Pro Tyr Ala Cys Pro Val Glu Ser
    330                 335                 340

tgt gat cgc cgc ttc tcc cgc tcc gac gag ctc acc cgc cac atc cgc
1350
Cys Asp Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg
345                 350                 355                 360

atc cac aca ggc cag aag ccc ttc cag tgc cgc atc tgc atg cgc aac
1398
Ile His Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn
                365                 370                 375

ttc agc cgc agc gac cac ctc acc acc cac atc cgc acc cac aca ggc
1446
Phe Ser Arg Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly
            380                 385                 390

gaa aag ccc ttc gcc tgc gac atc tgt gga aga aag ttt gcc agg agc
1494
Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser
        395                 400                 405

gat gaa cgc aag agg cat acc aag atc cac ttg cgg cag aag gac aag
1542
Asp Glu Arg Lys Arg His Thr Lys Ile His Leu Arg Gln Lys Asp Lys
    410                 415                 420

aaa gca gac aaa agt gtt gtg gcc tct tcg gcc acc tcc tct ctc tct
1590
Lys Ala Asp Lys Ser Val Val Ala Ser Ser Ala Thr Ser Ser Leu Ser
425                 430                 435                 440

tcc tac ccg tcc ccg gtt gct acc tct tac ccg tcc ccg gtt act acc
1638
Ser Tyr Pro Ser Pro Val Ala Thr Ser Tyr Pro Ser Pro Val Thr Thr
                445                 450                 455

tct tat cca tcc ccg gcc acc acc tca tac cca tcc cct gtg ccc acc
1686
Ser Tyr Pro Ser Pro Ala Thr Thr Ser Tyr Pro Ser Pro Val Pro Thr
            460                 465                 470

tcc ttc tcc tct ccc ggc tcc tcg acc tac cca tcc cct gtg cac agt
1734
Ser Phe Ser Ser Pro Gly Ser Ser Thr Tyr Pro Ser Pro Val His Ser
        475                 480                 485
```





```
2892

atttataaat atattcagga gttggaatgt tgtagttacc tactgagtag gcggcgattt
2952

ttgtatgtta tgaacatgca gttcattatt ttgtggttct attttacttt gtacttgtgt
3012

ttgcttaaac aaagtgactg tttggcttat aaacacattg aatgcgcttt attgcccatg
3072

ggatatgtgg tgtatatcct tccaaaaaat taaaacgaaa ataaagtagc tgcgattggg
3132
```

<210> 6
<211> 543
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Ala Ala Ala Lys Ala Glu Met Gln Leu Met Ser Pro Leu Gln Ile
1               5                  10                  15

Ser Asp Pro Phe Gly Ser Phe Pro His Ser Pro Thr Met Asp Asn Tyr
            20                  25                  30

Pro Lys Leu Glu Glu Met Met Leu Leu Ser Asn Gly Ala Pro Gln Phe
        35                  40                  45

Leu Gly Ala Ala Gly Ala Pro Glu Gly Ser Gly Ser Asn Ser Ser Ser
    50                  55                  60

Ser Ser Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly Ser Asn Ser Ser
65                  70                  75                  80

Ser Ser Ser Ser Thr Phe Asn Pro Gln Ala Asp Thr Gly Glu Gln Pro
                85                  90                  95

Tyr Glu His Leu Thr Ala Glu Ser Phe Pro Asp Ile Ser Leu Asn Asn
            100                 105                 110

Glu Lys Val Leu Val Glu Thr Ser Tyr Pro Ser Gln Thr Thr Arg Leu
        115                 120                 125

Pro Pro Ile Thr Tyr Thr Gly Arg Phe Ser Leu Glu Pro Ala Pro Asn
    130                 135                 140

Ser Gly Asn Thr Leu Trp Pro Glu Pro Leu Phe Ser Leu Val Ser Gly
145                 150                 155                 160

Leu Val Ser Met Thr Asn Pro Pro Ala Ser Ser Ser Ser Ala Pro Ser
                165                 170                 175
```

```
Ser Tyr Pro Ser Pro Val Thr Thr Ser Tyr Pro Ser Pro Ala Thr Thr
    450                 455                 460

Ser Tyr Pro Ser Pro Val Pro Thr Ser Phe Ser Ser Pro Gly Ser Ser
465                 470                 475                 480

Thr Tyr Pro Ser Pro Val His Ser Gly Phe Pro Ser Pro Ser Val Ala
                485                 490                 495

Thr Thr Tyr Ser Ser Val Pro Pro Ala Phe Pro Ala Gln Val Ser Ser
            500                 505                 510

Phe Pro Ser Ser Ala Val Thr Asn Ser Phe Ser Ala Ser Thr Gly Leu
        515                 520                 525

Ser Asp Met Thr Ala Thr Phe Ser Pro Arg Thr Ile Glu Ile Cys
    530                 535                 540
```

<210> 7
<211> 612
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (151)..(564)

<400> 7

```
gcatacctgc ctgggtgtcc aaatgtaact agatgctttc acaaacccca cccacaaagc
60

agcacatgtt tttaagactt cagttttcta ttcacatcgg cctcataata cccaccctga
120

cctgctgtaa aagacctgga acaaacaaaa atg att aca cct aca gtg aag atg
174
                                Met Ile Thr Pro Thr Val Lys Met
                                1               5

cac acc atg tcc tcc tcg cat ctc ttc tac ctg gcg ctg tgc ctg ctc
222
His Thr Met Ser Ser Ser His Leu Phe Tyr Leu Ala Leu Cys Leu Leu
    10                  15                  20

acc ttc acc agc tct gcc acg gct gga ccg gag acg ctc tgc ggg gct
270
Thr Phe Thr Ser Ser Ala Thr Ala Gly Pro Glu Thr Leu Cys Gly Ala
25                  30                  35                  40

gag ctg gtg gat gct ctt cag ttc gtg tgt gga gac agg ggc ttt tat
318
Glu Leu Val Asp Ala Leu Gln Phe Val Cys Gly Asp Arg Gly Phe Tyr
                45                  50                  55

ttc aac aag ccc aca ggg tat ggc tcc agc agt cgg agg gcg cct cag
366
Phe Asn Lys Pro Thr Gly Tyr Gly Ser Ser Ser Arg Arg Ala Pro Gln
            60                  65                  70
```





```
Ser Ala Gly Asn Lys Asn Tyr Arg Met
    130                 135
```

<210> 9
<211> 1038
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1038)

<400> 9

```
atg caa cgc ctg gtg gcc tgg gac cca gca tgt ctc ccc ctg ccg ccg
48
Met Gln Arg Leu Val Ala Trp Asp Pro Ala Cys Leu Pro Leu Pro Pro
1               5                   10                  15

ccg ccg cct gcc ttt aaa tcc atg gaa gtg gcc aac ttc tac tac gag
96
Pro Pro Pro Ala Phe Lys Ser Met Glu Val Ala Asn Phe Tyr Tyr Glu
            20                  25                  30

gcg gac tgc ttg gct gct gcg tac ggc ggc aag gcg gcc ccc gcg gcg
144
Ala Asp Cys Leu Ala Ala Ala Tyr Gly Gly Lys Ala Ala Pro Ala Ala
        35                  40                  45

ccc ccc gcg gcc aga ccc ggg ccg cgc ccc ccc gcc ggc gag ctg ggc
192
Pro Pro Ala Ala Arg Pro Gly Pro Arg Pro Pro Ala Gly Glu Leu Gly
    50                  55                  60

agc atc ggc gac cac gag cgc gcc atc gac ttc agc ccg tac ctg gag
240
Ser Ile Gly Asp His Glu Arg Ala Ile Asp Phe Ser Pro Tyr Leu Glu
65                  70                  75                  80

ccg ctg ggc gcg ccg cag gcc ccg gcg ccc gcc acg gcc acg gac acc
288
Pro Leu Gly Ala Pro Gln Ala Pro Ala Pro Ala Thr Ala Thr Asp Thr
                85                  90                  95

ttc gag gcg gct ccg ccc gcg ccc gcc ccc gcg ccc gcc tcc tcc ggg
336
Phe Glu Ala Ala Pro Pro Ala Pro Ala Pro Ala Pro Ala Ser Ser Gly
            100                 105                 110

cag cac cac gac ttc ctc tcc gac ctc ttc tcc gac gac tac ggg ggc
384
Gln His His Asp Phe Leu Ser Asp Leu Phe Ser Asp Asp Tyr Gly Gly
        115                 120                 125

aag aac tgc aag aag ccg gcc gag tac ggc tac gtg agc ctg ggg cgc
432
Lys Asn Cys Lys Lys Pro Ala Glu Tyr Gly Tyr Val Ser Leu Gly Arg
    130                 135                 140

ctg ggg gct gcc aag ggc gcg ctg cac ccc ggc tgc ttc gcg ccc ctg
480
Leu Gly Ala Ala Lys Gly Ala Leu His Pro Gly Cys Phe Ala Pro Leu
145                 150                 155                 160

cac cca ccg ccc ccg ccg ccg ccg ccg ccc gcc gag ctc aag gcg gag
528
```

```
Met Gln Arg Leu Val Ala Trp Asp Pro Ala Cys Leu Pro Leu Pro Pro
1               5                  10                  15

Pro Pro Pro Ala Phe Lys Ser Met Glu Val Ala Asn Phe Tyr Tyr Glu
            20                  25                  30

Ala Asp Cys Leu Ala Ala Ala Tyr Gly Gly Lys Ala Ala Pro Ala Ala
        35                  40                  45

Pro Pro Ala Ala Arg Pro Gly Pro Arg Pro Pro Ala Gly Glu Leu Gly
    50                  55                  60

Ser Ile Gly Asp His Glu Arg Ala Ile Asp Phe Ser Pro Tyr Leu Glu
65                  70                  75                  80

Pro Leu Gly Ala Pro Gln Ala Pro Ala Pro Ala Thr Ala Thr Asp Thr
                85                  90                  95

Phe Glu Ala Ala Pro Pro Ala Pro Ala Pro Ala Pro Ala Ser Ser Gly
            100                 105                 110

Gln His His Asp Phe Leu Ser Asp Leu Phe Ser Asp Asp Tyr Gly Gly
        115                 120                 125

Lys Asn Cys Lys Lys Pro Ala Glu Tyr Gly Tyr Val Ser Leu Gly Arg
    130                 135                 140

Leu Gly Ala Ala Lys Gly Ala Leu His Pro Gly Cys Phe Ala Pro Leu
145                 150                 155                 160

His Pro Pro Pro Pro Pro Pro Pro Pro Pro Ala Glu Leu Lys Ala Glu
                165                 170                 175

Pro Gly Phe Glu Pro Ala Asp Cys Lys Arg Lys Glu Glu Ala Gly Ala
            180                 185                 190

Pro Gly Gly Gly Ala Gly Met Ala Ala Gly Phe Pro Tyr Ala Leu Arg
        195                 200                 205

Ala Tyr Leu Gly Tyr Gln Ala Val Pro Ser Gly Ser Ser Gly Ser Leu
    210                 215                 220

Ser Thr Ser Ser Ser Ser Ser Pro Pro Gly Thr Pro Ser Pro Ala Asp
225                 230                 235                 240

Ala Lys Ala Pro Pro Thr Ala Cys Tyr Ala Gly Ala Gly Pro Ala Pro
                245                 250                 255

Ser Gln Val Lys Ser Lys Ala Lys Lys Thr Val Asp Lys His Ser Asp
            260                 265                 270
```





```
720

tcc atg aaa gcc ctc aga ttg tcg gct tcc gcc ctc ttc tgc ctt ctg
768
    Met Lys Ala Leu Arg Leu Ser Ala Ser Ala Leu Phe Cys Leu Leu
    1               5                  10                  15

ctg atc aac ggg tta ggg gca gca ccc cct ggt cgc cct gag gcg cag
816
Leu Ile Asn Gly Leu Gly Ala Ala Pro Pro Gly Arg Pro Glu Ala Gln
                20                  25                  30

cct cct cct ctc agc tct gag cat aaa gag ccg gta gcc ggg gac gca
864
Pro Pro Pro Leu Ser Ser Glu His Lys Glu Pro Val Ala Gly Asp Ala
            35                  40                  45

gtg ccc ggg cca aag gat ggc agc gcc cca gag gtc cga ggc gct cgg
912
Val Pro Gly Pro Lys Asp Gly Ser Ala Pro Glu Val Arg Gly Ala Arg
        50                  55                  60

aat tcc gag ccg cag gac gag gga gag ctt ttc cag ggc gtg gat ccc
960
Asn Ser Glu Pro Gln Asp Glu Gly Glu Leu Phe Gln Gly Val Asp Pro
    65                  70                  75

cgg gcg ctg gcc gcg gtg ctg ctg cag gca ctc gac cgt ccc gcc tca
1008
Arg Ala Leu Ala Ala Val Leu Leu Gln Ala Leu Asp Arg Pro Ala Ser
80                  85                  90                  95

ccc ccg gca cca agc ggc tcc cag cag ggg ccg gag gaa gaa gca gct
1056
Pro Pro Ala Pro Ser Gly Ser Gln Gln Gly Pro Glu Glu Glu Ala Ala
                100                 105                 110

gaa gct ctg ctg acc gag acc gtg cgc agc cag acc cac agc ctc ccg
1104
Glu Ala Leu Leu Thr Glu Thr Val Arg Ser Gln Thr His Ser Leu Pro
            115                 120                 125

gcg gcc gga gag ccc gag ccc gcg gcg ccc cct cgc cct cag act ccg
1152
Ala Ala Gly Glu Pro Glu Pro Ala Ala Pro Pro Arg Pro Gln Thr Pro
        130                 135                 140

gag aat ggg ccc gag gcg agc gat ccc tcc gag gag ctc gag gcg cta
1200
Glu Asn Gly Pro Glu Ala Ser Asp Pro Ser Glu Glu Leu Glu Ala Leu
    145                 150                 155

gcg tcc ctg ctc cag gaa ctg cga gat ttc agt cca agt agc gcc aag
1248
Ala Ser Leu Leu Gln Glu Leu Arg Asp Phe Ser Pro Ser Ser Ala Lys
160                 165                 170                 175

cgc cag cag gag acg gcg gca gca gag acg gaa acc cgc acg cac acg
1296
Arg Gln Gln Glu Thr Ala Ala Ala Glu Thr Glu Thr Arg Thr His Thr
                180                 185                 190

ctg acc cga gtg aat ctg gag agc ccg ggg cca gag cgc gta tgg cgc
1344
Leu Thr Arg Val Asn Leu Glu Ser Pro Gly Pro Glu Arg Val Trp Arg
            195                 200                 205
```

```
gcc gag gac aag cgc tcc cag gag gag acg ccg ggc cac cgg cgg aag
2016
Ala Glu Asp Lys Arg Ser Gln Glu Glu Thr Pro Gly His Arg Arg Lys
                420                 425                 430

gag gcc gag ggg aca gag gag ggc ggg gag gag gag gac gac gag gag
2064
Glu Ala Glu Gly Thr Glu Glu Gly Gly Glu Glu Glu Asp Asp Glu Glu
            435                 440                 445

atg gat ccg cag acg atc gac agc ctc att gag ctg tcc acc aaa ctc
2112
Met Asp Pro Gln Thr Ile Asp Ser Leu Ile Glu Leu Ser Thr Lys Leu
        450                 455                 460

cac ctg cca gcg gac gac gtg gtc agc atc atc gag gag gtg gag gag
2160
His Leu Pro Ala Asp Asp Val Val Ser Ile Ile Glu Glu Val Glu Glu
    465                 470                 475

aag cgg aac cga aag aag aaa gcc cct ccc gag ccc gtg ccg ccc ccc
2208
Lys Arg Asn Arg Lys Lys Lys Ala Pro Pro Glu Pro Val Pro Pro Pro
480                 485                 490                 495

cgt gcc gcc ccc gcc ccc acc cac gtc cgc tcc ccg cag ccc ccg ccc
2256
Arg Ala Ala Pro Ala Pro Thr His Val Arg Ser Pro Gln Pro Pro Pro
                500                 505                 510

ccg ccc ccg tcc gca cga gac gag ctg ccg gac tgg aac gag gtg ctc
2304
Pro Pro Pro Ser Ala Arg Asp Glu Leu Pro Asp Trp Asn Glu Val Leu
            515                 520                 525

ccg ccc tgg gat cgg gag gag gac gag gtg tac ccg cca ggg ccg tac
2352
Pro Pro Trp Asp Arg Glu Glu Asp Glu Val Tyr Pro Pro Gly Pro Tyr
        530                 535                 540

cac cct ttc ccc aac tac atc cgg ccg cgg aca ctg cag ccg ccc tcg
2400
His Pro Phe Pro Asn Tyr Ile Arg Pro Arg Thr Leu Gln Pro Pro Ser
    545                 550                 555

gcc ttg cgc cgc cgc cac tac cac cac gcc ttg ccg cct tcg cgc cac
2448
Ala Leu Arg Arg Arg His Tyr His His Ala Leu Pro Pro Ser Arg His
560                 565                 570                 575

tat ccc ggc cgg gag gcc cag gcc cgg cac gcg cag cag gag gag gcg
2496
Tyr Pro Gly Arg Glu Ala Gln Ala Arg His Ala Gln Gln Glu Glu Ala
                580                 585                 590

gag gcg gag gag cgc cgg ctg cag gag cag gag gag ctg gag aat tac
2544
Glu Ala Glu Glu Arg Arg Leu Gln Glu Gln Glu Glu Leu Glu Asn Tyr
            595                 600                 605

atc gag cac gtg ctg ctc cgg cgc ccg tga ctgcccttcc cggtcccgcc
2594
Ile Glu His Val Leu Leu Arg Arg Pro
        610                 615

cccgcgcgcc cccgccgcgc gcgcgcgccg gcgccccccc ccgtgttgct ccccctcggt
```





```
Asn Gly Pro Glu Ala Ser Asp Pro Ser Glu Glu Leu Glu Ala Leu Ala
145                 150                 155                 160

Ser Leu Leu Gln Glu Leu Arg Asp Phe Ser Pro Ser Ser Ala Lys Arg
                165                 170                 175

Gln Gln Glu Thr Ala Ala Ala Glu Thr Glu Thr Arg Thr His Thr Leu
            180                 185                 190

Thr Arg Val Asn Leu Glu Ser Pro Gly Pro Glu Arg Val Trp Arg Ala
        195                 200                 205

Ser Trp Gly Glu Phe Gln Ala Arg Val Pro Glu Arg Ala Pro Leu Pro
    210                 215                 220

Pro Pro Ala Pro Ser Gln Phe Gln Ala Arg Met Pro Asp Ser Gly Pro
225                 230                 235                 240

Leu Pro Glu Thr His Lys Phe Gly Glu Gly Val Ser Ser Pro Lys Thr
                245                 250                 255

His Leu Gly Glu Ala Leu Ala Pro Leu Ser Lys Ala Tyr Gln Gly Val
            260                 265                 270

Ala Ala Pro Phe Pro Lys Ala Arg Arg Ala Glu Ser Ala Leu Leu Gly
        275                 280                 285

Gly Ser Glu Ala Gly Glu Arg Leu Leu Gln Gln Gly Leu Ala Gln Val
    290                 295                 300

Glu Ala Gly Arg Arg Gln Ala Glu Ala Thr Arg Gln Ala Ala Ala Gln
305                 310                 315                 320

Glu Glu Arg Leu Ala Asp Leu Ala Ser Asp Leu Leu Leu Gln Tyr Leu
                325                 330                 335

Leu Gln Gly Gly Ala Arg Gln Arg Gly Leu Gly Gly Arg Gly Leu Gln
            340                 345                 350

Glu Ala Ala Glu Glu Arg Glu Ser Ala Arg Glu Glu Glu Glu Ala Glu
        355                 360                 365

Gln Glu Arg Arg Gly Gly Glu Glu Arg Val Gly Glu Glu Asp Glu Glu
    370                 375                 380

Ala Ala Glu Ala Ala Glu Ala Glu Ala Asp Glu Ala Glu Arg Ala Arg
385                 390                 395                 400

Gln Asn Ala Leu Leu Phe Ala Glu Glu Glu Asp Gly Glu Ala Gly Ala
```